NCVER

TECHNICAL PAPER

Estimating apprentice and trainee completion and attrition rates using a 'life tables' approach

TOM KARMEL
PETER MLOTKOWSKI

NATIONAL CENTRE FOR VOCATIONAL EDUCATION RESEARCH

FOR CLARITY WITH FIGURES,
PLEASE PRINT IN COLOUR

Australian Government
Department of Education, Employment and Workplace Relations

TD/TNC 104.09

Published by NCVER
ABN 87 007 967 311

Level 11, 33 King William Street, Adelaide, SA 5000
PO Box 8288 Station Arcade, Adelaide SA 5000, Australia

ph +61 8 8230 8400 fax +61 8 8212 3436
email ncver@ncver.edu.au
<http://www.ncver.edu.au>
<http://www.ncver.edu.au/publications/2389.html>

About the research

NCVER

Estimating apprentice and trainee completion and attrition rates using a 'life tables' approach

Tom Karmel and Peter Mlotkowski, NCVER

The proportion of apprentices and trainees completing their contract of training has received considerable attention in recent years. To support public discussion, the National Centre for Vocational Education Research (NCVER) has been publishing tabulations of both completion and attrition rates. Up until recently, the approach taken was to identify a commencing cohort and follow it over time. Currently, NCVER publishes completion and attrition rates that are calculated from tabulations of completions and cancellations/withdrawals within particular quarters, for particular commencing cohorts (see NCVER 2010). One drawback of both approaches is that the resultant completion and attrition rates refer to cohorts which commenced some time earlier.

The economic downturn provided the impetus for the development of completion and attrition rates for the latest commencing cohorts. Such up-to-date estimates are possible if they are modelled using recent cross-sectional data to mimic data with a time dimension. This is the basic approach used in the construction of 'life tables'.

This technical paper details a methodological approach which constructs cross-sectional estimates of completion and attrition rates for the latest commencing cohorts. These reflect what the completion and attrition rates would be if the patterns observed in a particular quarter remained unchanged for the life of the commencing cohort in that quarter. Throughout the paper, we use a simple and contrived example to illustrate the model, along with the formal mathematical notation. Finally, we apply the model to cohorts commencing in the December quarters of 2007 to 2009.

Readers who would like to reproduce the estimates provided in this technical paper are directed to the support document in Excel format.

Tom Karmel
Managing Director, NCVER

Contents

Tables and figures

Tables

Figures

Introduction

The proportion of apprentices and trainees completing their contract of training has received considerable attention in recent years. To support public discussion, the National Centre for Vocational Education Research (NCVER) has been publishing tabulations of both completion rates and attrition rates in its annual apprenticeship and traineeship statistical bulletin (NCVER 2010). Both measures are closely related, since by definition, once the contract of training has expired, an individual must have either completed or not. The advantage of an attrition rate is that it has a time dimension, so we can say what proportion of apprentices or trainees have withdrawn (or had their contract cancelled) within three months of commencement, six months and so on. While one could calculate a progressive completion rate in the same way, this makes less sense because completion typically involves a predetermined amount of time.

The approach taken by NCVER in the past was to identify a commencing cohort and follow it over time. One drawback of this method was that the resultant completion and attrition rates referred to cohorts which commenced quite some time earlier. This is because there are lags in the data, which mean that completions and cancellations/withdrawals are estimated for recent quarters. While these estimates are fine for aggregates, they must be discarded for the cohort analysis that matches individual contracts over time. For this reason, NCVER has changed its approach and now calculates completion and attrition rates from tabulations of completions and cancellations/withdrawals within particular quarters, for particular commencing cohorts (see NCVER 2010). However, because of the durations involved, the resultant rates will never refer to the very latest commencing cohorts. A traditional trade apprenticeship typically takes four years to complete, while a non-trade traineeship is shorter, around two years for the higher levels.

The economic downturn provided the impetus for the development of completion and attrition rates for the latest commencing cohorts. In one sense, arriving at such up-to-date estimates are impossible, because we will not observe the outcomes of the latest commencing cohorts for some years. However, up-to-date estimates are possible if they are modelled using recent cross-sectional data to mimic data with a time dimension. This is the basic approach used in the construction of 'life tables'.

The purpose of this paper is to develop a model which does this. The next section develops the methodology, beginning with the calculation of what we term 'transition probabilities'. These simply compare the number of apprentices and trainees who completed or were cancelled/withdrawn in a quarter with the number surviving to the previous quarter, conditional on the period of time since commencement of the apprenticeship or traineeship. The crux of the method is the assumption that the latest commencing cohorts will behave according to the cross-sectional transitions derived. We use a simple and contrived example to illustrate this, along with the formal mathematical notation.

One drawback of our approach is that it inherently assumes that all commencements occur at a point in time, specifically the beginning of each quarter, rather than allowing for commencements to occur over the whole quarter. However, taking this factor into account greatly complicates the estimation of the transitions. One way around the issue is to simply assume that all commencements in the first quarter begin at exactly half way through the quarter, so that the resultant completion and attrition rates refer to 0.5 quarter, 1.5 quarters, 2.5 quarters and so on.

Before presenting completion and attrition rates for the latest commencing cohorts, we test how well the 'life tables' approach works when applied to historical apprenticeship and traineeship data (the third section). To do this we take a cohort which commenced quite some time earlier, the September quarter 2005, and calculate actual completion and attrition rates from the latest data available. We then match these with the cross-sectional rates implied by the model, using data as they were first reported for that cohort. The results show that the two are remarkably similar. For completion, the cross-sectional rates follow the actual rates closely before diverging slightly in later quarters. For attrition, there is virtually no difference between the two for the whole time period.

While these results are reassuring, we emphasise that the cross-section estimates are based on the quarter-to-quarter transition probabilities, for various periods since commencement, as observed at a specific quarter. Thus the cross-sectional estimates can be interpreted as a completion rate based on the information available at the time point in question rather than as an estimate of a rate that has not yet been observed. Under this interpretation, the cross-sectional estimates are legitimate statistics in their own right, even if future transition patterns change. This interpretation leads naturally to comparisons between the cross-sectional estimates across years.

Finally, we present cross-sectional completion and attrition rates for the latest commencing trade and non-trade cohorts. We choose the cohorts commencing in the December quarters of 2007 to 2009. These periods cover the economic cycle in Australia and we are interested in the differences in the patterns for completion and attrition between the three cohorts.

The results show that, with deteriorating economic conditions from the end of 2007 to the end of 2009, completion rates have increased and attrition rates have decreased, with the differences being quite dramatic in the non-trades.

The paper ends with some concluding comments.

Methodology

Calculating ‘transition probabilities’

The first step is to calculate what we term ‘transition probabilities’. These are based on simple tabulations of completions and cancellations/withdrawals within particular quarters, for particular commencing cohorts.

We begin by identifying three states: in-training or continuing; completed; and cancelled/withdrawn. An individual commences a contract of training and from that point on he/she must be in one of these three states. The two states of completed and cancelled/withdrawn are absorbing states: once an individual has completed or cancelled/withdrawn, then the person can no longer have a continuing contract. This means that eventually every contract of training must result in either a completion or a cancellation/withdrawal.

Consider the very simple and contrived example in table 1; we will provide the formal mathematical notation later. For simplicity, there are only four quarters and each sees 100 contracts commenced. We observe the data at quarter four, meaning that only for the cohort commencing at quarter one have all contracts resulted in either a completion or a cancellation/withdrawal. Consider this cohort more closely. In the first quarter, five contracts were completed and 20 were cancelled/withdrawn, to leave 75 continuing. In the next quarter, a further ten contracts were completed and 15 were cancelled/withdrawn, to leave 50 continuing. By the fourth quarter, every contract has resulted in either a completion or a cancellation/ withdrawal. For simplicity, these patterns are the same for each of the commencing cohorts.

Table 1 Number of completed, cancelled/withdrawn and continuing contracts of training by commencing quarter

Commenced		Completed				Cancelled/withdrawn				Continuing			
		Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Q1	100	5	10	15	20	20	15	10	5	75	50	25	0
Q2	100		5	10	15		20	15	10		75	50	25
Q3	100			5	10			20	15			75	50
Q4	100				5				20				75

We can use these tabulations of completions and cancellations/withdrawals to calculate the transition probabilities (table 2).

Table 2 **Completed, cancelled/withdrawn and continuing transition probabilities by commencing quarter**

Commenced	Completed				Cancelled/withdrawn				Continuing			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Q1	0.05	0.13	0.30	0.80	0.20	0.20	0.20	0.20	0.75	0.67	0.50	0.00
Q2		0.05	0.13	0.30		0.20	0.20	0.20		0.75	0.67	0.50
Q3			0.05	0.13			0.20	0.20			0.75	0.67
Q4				0.05				0.20				0.75

The table is best explained by considering an example: the cohort commencing quarter one. In the first quarter, five contracts out of the commencing 100 were completed to give a transition probability of 0.05. In the next quarter, ten contracts out of the 75 continuing from the previous quarter were completed to give a rate of 0.13, and so on, until the fourth quarter (0.80 = 20/25). The cancellation/withdrawal rates follow the same logic. Put simply, the transition probabilities in table 2 compare the number of contracts that were completed or cancelled/withdrawn in a quarter with the number that survived to the previous quarter. The continuing rates are derived differently—one minus the completion rate minus the cancellation/withdrawal rate—something which follows from the assumption that a contract must be in one of the three states of continuing, completed or cancelled/withdrawn.

Table 2 is a transition matrix and may be defined formally. Consider a cohort of contracts commencing at time *t* (measured in quarters), which denotes the number of quarters since a given start time. Denote the number of contracts in this cohort continuing during quarter *t+i -1* as $N_{t,i}^{c}$.

Note that $i = 1$ is the first quarter for each cohort and $N_{t,t}^{c}$ is taken to be the number of commencements in the quarter *t*.

We also define the number of completed contracts in quarter *t+i-1* (of those commencing in quarter *t*) as $N_{t,i}^{f}$ and the number of cancellations/withdrawals in quarter *t+i-1* (of those commencing in quarter t) as $N_{t,i}^{w}$. Clearly we need to have:

$$N_{t,i+1}^{c} = N_{t,i}^{c} - N_{t,i}^{f} - N_{t,i}^{w}$$

Let us consider a single cohort, and suppress the index *t*. Then the transition matrix for quarter *i* is given by:

$$T_i = \begin{bmatrix} 1 - N_i^f / N_{i-1}^c - N_i^w / N_{i-1}^c & N_i^f / N_{i-1}^c & N_i^w / N_{i-1}^c \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

with the second and third columns expressing the transition probabilities for completion and cancellation/withdrawal, respectively.

Applying the 'life tables' approach

In our example, the latest cohort commenced in quarter four. We now wish to find out what proportion of these contracts will be completed and what proportion will be cancelled/withdrawn. In one sense, this is not possible because we will not observe the final outcome of this cohort for another three quarters—using a more realistic example, we would not observe the final outcome of the latest commencing cohort for several years. However, such up-to-date estimates are possible if we assume that the cohort will behave according to the transition probabilities constructed in table 2. That is, we will use recent cross-sectional data to mimic data with a time dimension. This is the basic approach used in the construction of 'life tables'.

We continue our example. The cohort we are interested in is the one which commenced at quarter four. Table 3 presents the cross-sectional transition probabilities for this cohort. Notice that this table is merely an abridged version of table 2, with only the transition probabilities for quarter four retained for each state.

Table 3 Cross-sectional transition probabilities for cohort commencing quarter four

Commenced	Completed	Cancelled/withdrawn	Continuing
	Q4	Q4	Q4
Q1	0.80	0.20	0.00
Q2	0.30	0.20	0.50
Q3	0.13	0.20	0.67
Q4	0.05	0.20	0.75

From table 3 we can clearly see that the transition probabilities for each state are based on data going from one quarter to the next, the implication of this being that each rate is based on a different commencing cohort. That is, early transition probabilities (0.05, 0.20 and 0.75 for completed, cancelled/withdrawn and continuing, respectively) are based on the most recent commencements, while later probabilities (0.80, 0.20 and 0.00, for completed, cancelled/ withdrawn and continuing, respectively) are based on cohorts which commenced earlier.

In this sense, table 3 needs to be 'flipped over' to show how the cohort commencing in quarter four will behave (table 4).

Table 4 'Flipped' cross-sectional transition probabilities for cohort commencing quarter four

Time period	Completed	Cancelled/withdrawn	Continuing
1	0.05	0.20	0.75
2	0.13	0.20	0.67
3	0.30	0.20	0.50
4	0.80	0.20	0.00

Notice that we have not performed any additional calculations in the last two steps. We have merely taken the transition probabilities in table 2, and for a particular cohort, 'flipped' them over. The only difficult part of the exercise is recognising that the latest commencing cohort will behave according to transition probabilities calculated for different commencing cohorts, with the transitions from the earliest commencing cohorts being used for the later quarters.

It is a simple step to now calculate the cumulative rates (table 5).

Table 5 Cumulative cross-sectional rates for cohort commencing quarter four

Quarter ending	Completed	Cancelled/withdrawn	Continuing
0.0	0.00	0.00	1.00
0.5	0.05	0.20	0.75
1.5	0.15	0.35	0.50
2.5	0.30	0.45	0.25
3.5	0.50	0.50	0.00

The time periods used are half a quarter for the first time period and one quarter thereafter. That is, we assume that all commencements in the first quarter begin at exactly half way through the quarter rather than at the beginning. In reality, commencements occur over the whole quarter rather than at any point in time, but as already highlighted, the requisite calculations complicate matters for little gain in precision.

For completeness, let us explain how the cumulative rates are calculated. At the beginning, every contract is continuing. At the quarter ending 0.5, the cumulative rates equal the transition probabilities for time period one in table 4. At the quarter ending 1.5, the cumulative completion rate is 0.15 (= 0.75*0.13+0.05) and the cumulative attrition rate is 0.35 (= 0.75*0.20+0.20). At the quarter ending 3.5, exactly half of the contracts have resulted in a completion and half have been cancelled/withdrawn.

We are able to define the steps formally. Define π_1, α_1 and β_1 as the transition probabilities for the first quarter of a contract, where the three states are as before: continuing; completed; and cancelled/withdrawn. That is, we have a transition matrix for the first quarter:

$$\begin{bmatrix} \pi_1 & \alpha_1 & \beta_1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Similarly, we have a transition matrix for quarter *i*:

$$\begin{bmatrix} \pi_i & \alpha_i & \beta_i \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

By multiplying the transition matrices, we can describe the cumulative completion and attrition rates. We commence with an initial vector (1,0,0). Denote Ω_i as the transition matrix for the i^{th} time period. Then the state vector $\begin{bmatrix} p_1^t & p_2^t & p_3^t \end{bmatrix}$ at time t is given by:

$$\begin{bmatrix} p_1^t & p_2^t & p_3^t \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \Omega_1 \Omega_2 \ldots \Omega_t$$

The three elements in the state vector represent the cumulative proportion of continuing, completed and cancelled/withdrawn contracts after t time periods.

Application

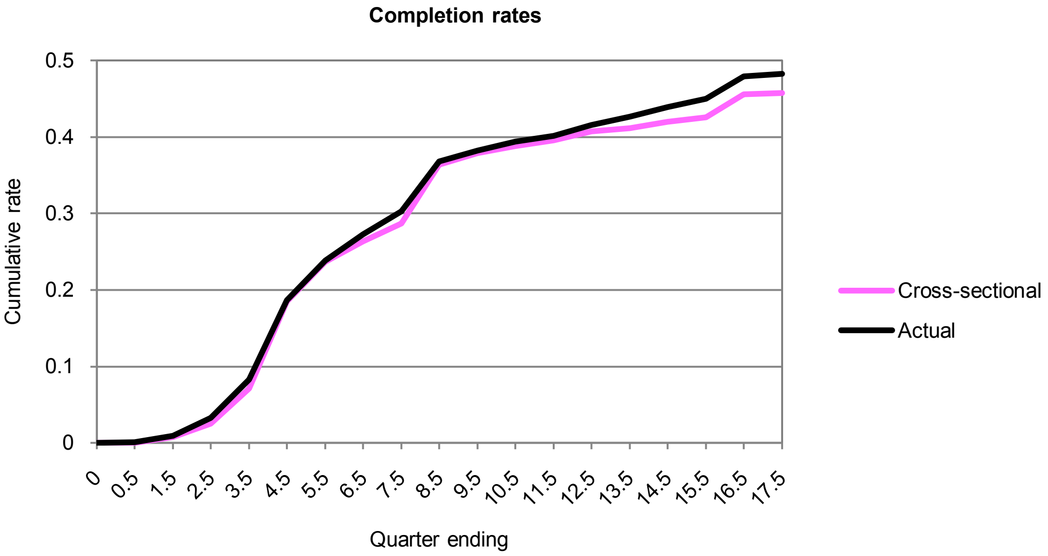
How well does the 'life tables' approach work?

Before presenting completion and attrition rates for the latest commencing cohorts, it is a good idea to test how well the 'life tables' approach works when applied to apprenticeship and traineeship data.

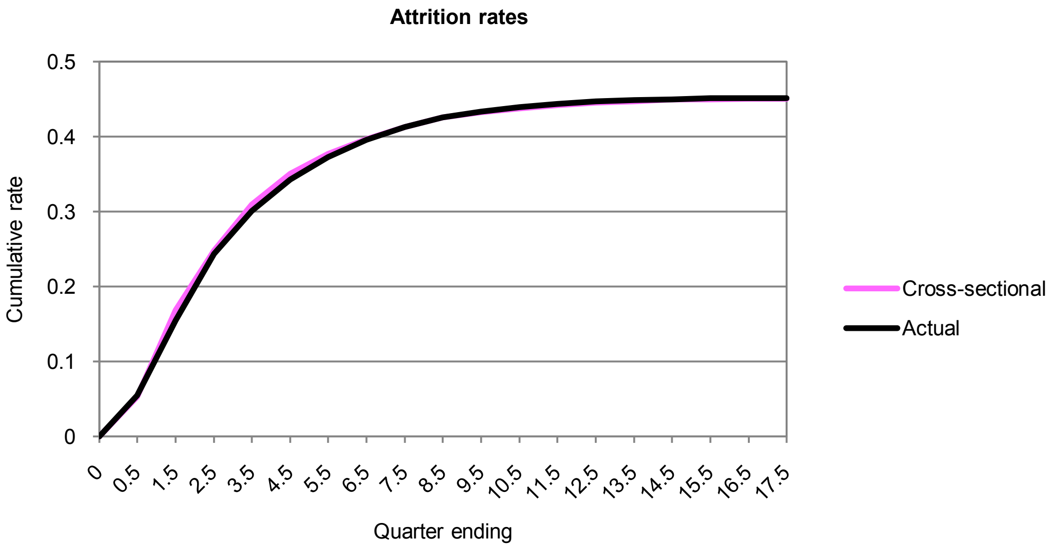
To do this, we consider a cohort which commenced quite some time earlier, say, the September quarter 2005. A sufficient period of time has passed for us to calculate actual completion and attrition rates for this cohort from the latest data available. We can also go back in time and look at the data as they were first reported for this cohort and use the model developed in the previous section to calculate the cross-sectional completion and attrition rates.[1]

Figure 1 presents the results of this exercise. At this stage we do not disaggregate into trade/non-trade contracts of training—we are interested in the robustness of our method in an overall sense. The results show that the actual and cross-sectional rates are remarkably similar. For completion, the cross-sectional rates follow the actual rates closely before diverging slightly in later quarters. For attrition, there is virtually no difference between the two for the whole time period.

Figure 1 Cross-sectional and actual completion and attrition rates for cohort commencing September quarter, 2005



[1] Alternatively, we could simply use the latest data to derive both sets of rates. However, this would be a less robust because of the nature of the data collection, and the fact that activity levels may be revised for a period of up to two years. We are interested in comparing rates as they would have been calculated 'back then' with what actually happened.

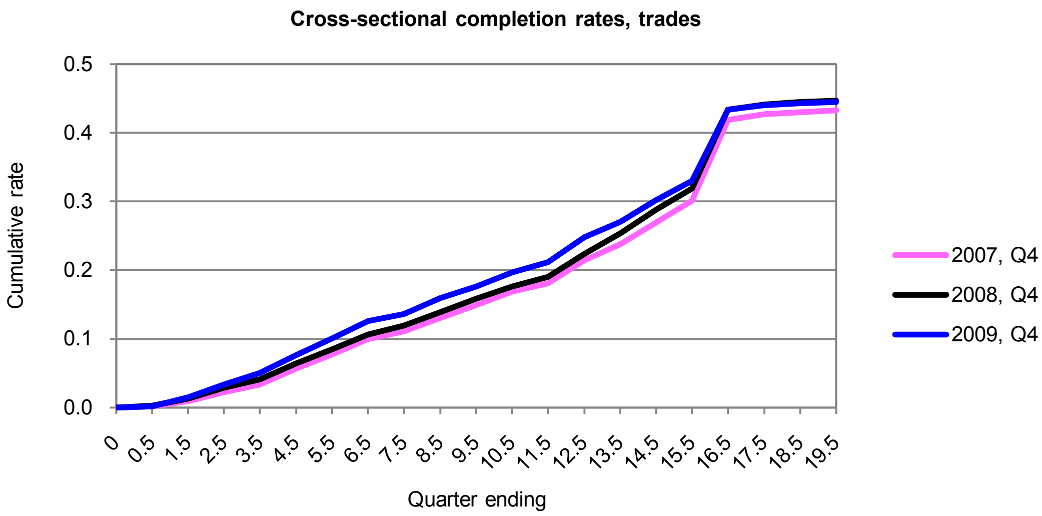


Source: Derived from the National Apprentice and Trainee Collection, no. 63 and 46.

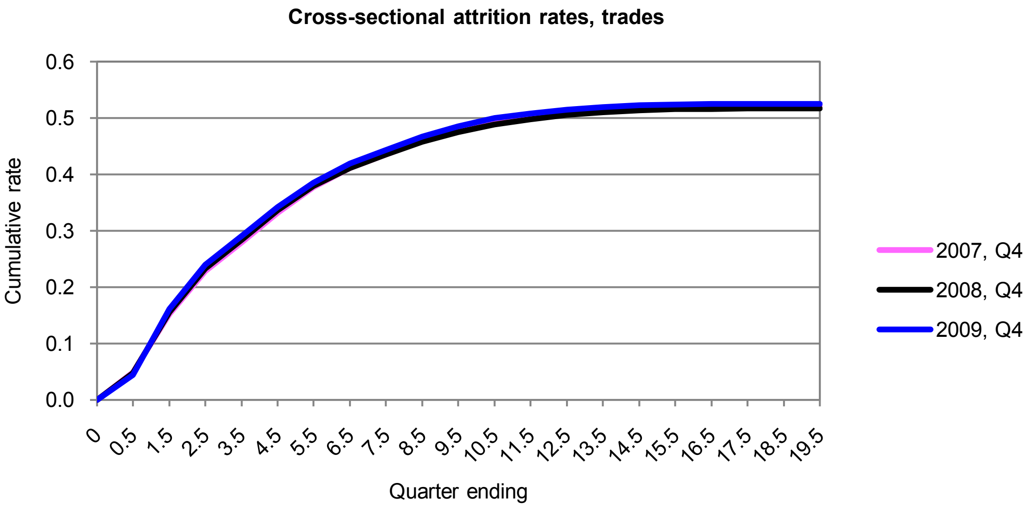
Estimates for latest commencing cohorts

We present estimates of cross-sectional completion and attrition rates for cohorts commencing in the December quarters of 2007 to 2009. These periods cover the economic cycle in Australia, with the unemployment rate peaking at 5.8% in October 2009, from an historical low of 4.0% in March 2008 (ABS 2010). We are interested in the differences in the patterns for completion and attrition between the three cohorts. Noting that the economic cycle affects apprenticeships very differently from traineeships, we provide separate estimates for trade and non-trade contracts of training.[2] The underlying calculations behind the estimates may be found in the supporting document in Excel format.

Figure 2 Cross-sectional completion and attrition rates for cohorts commencing December quarter, 2007–09, trades



[2] See Karmel and Misko (2009) for a discussion on the impact of economic downturns on apprenticeships and traineeships.

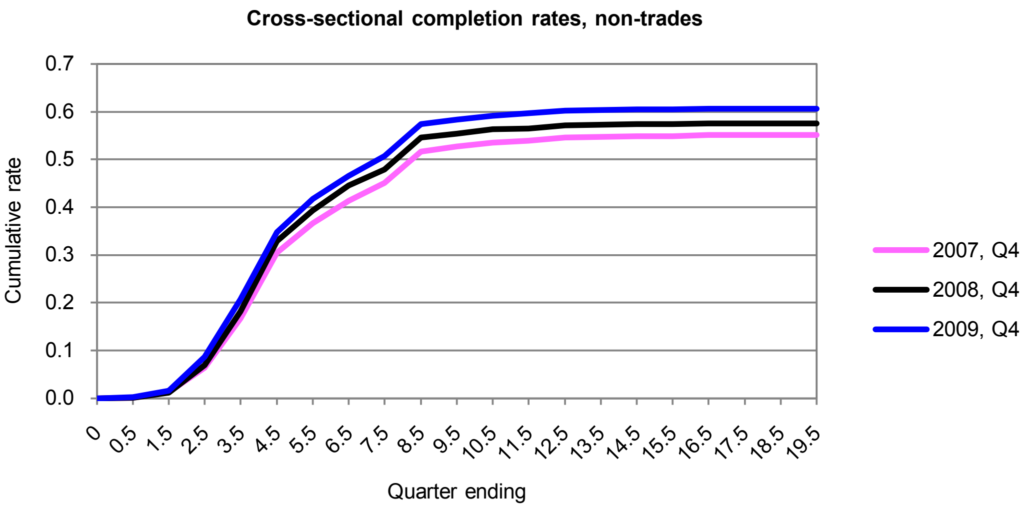


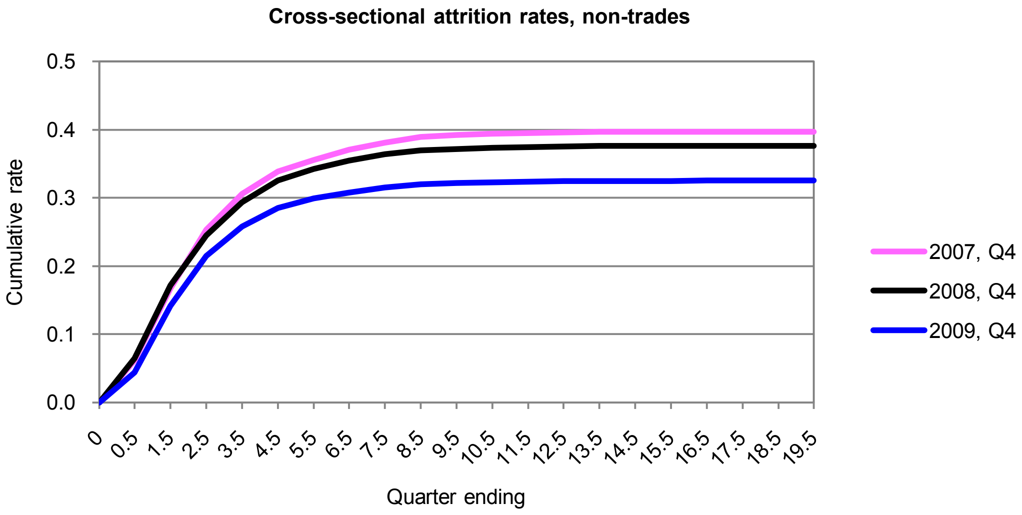
Source: Derived from the National Apprentice and Trainee Collection, no. 63.

The cumulative completion and attrition rates look quite different. The attrition rates are smooth, reflecting cancellations and withdrawals occurring at all durations but at a faster rate in early quarters. By contrast, the completion rates start off slowly (very short trade apprenticeships are a rarity) and kick up after around 16 quarters, which historically has been the standard period for a trade apprenticeship. Our interest lies in the differences between the patterns for the three cohorts. Overall, completion rates have increased slightly from the end of 2007 to the end of 2009, while attrition rates are virtually the same. Although the overall rate has increased only slightly, completions have clearly tended to occur faster with deteriorating economic conditions.

Analogous graphs for trainees are presented in figure 3.

Figure 3 Cross-sectional completion and attrition rates for cohorts commencing December quarter, 2007–09, non-trades





Source: Derived from the National Apprentice and Trainee Collection, no. 63.

For trainees, the differences between the three cohorts are far more dramatic. Completion rates have clearly increased with deteriorating economic conditions from the end of 2007 to the end of 2009, while attrition rates have decreased markedly, particularly for the cohort commencing in December 2009.

Concluding comments

In this short paper we have set up a framework for estimating completion and attrition rates for the latest commencing cohorts using a 'life tables' methodology and simple cross-tabulations of commencements, completions and cancellations/withdrawals. These estimates reflect what the completion and attrition rates would be had the patterns observed in a particular quarter remained unchanged for the life of the commencing cohort in that quarter.

One drawback of our approach is the boundary effect caused by commencements occurring across a quarter rather than at a point in time. Our way around this is to simply assume that the commencements occur exactly half way through the quarter. Our approach does have the advantage of simplicity, and we have shown that, for the cohort commencing September quarter 2005, the cross-sectional completion and attrition rates follow actual rates very closely.

We presented estimates of cross-sectional completion and attrition rates for cohorts commencing in the December quarters of 2007 to 2009. These periods cover the economic cycle in Australia. The results show that from the end of 2007 to the end of 2009, completion rates have increased and attrition rates have decreased, with the differences being quite dramatic in the non-trades.

References

ABS (Australian Bureau of Statistics) 2010, *Labour force, Australia, February 2010,* cat.no.6202.0, ABS, Canberra.

Karmel, T & Misko, J 2009, *Apprenticeships and traineeships in the downturn*, NCVER, Adelaide.

NCVER (National Centre for Vocational Education Research) 2010, *Australian vocational education and training statistics: Apprentices and trainees 2009 – Annual*, NCVER, Adelaide.